FIN 3370-Research Methods in Finance

Name: Zhiming Mei
Student ID: 120090007
Assignment Title: Assignment 1
Course Instructor: Prof. Zongbo Huang

Contents

1 Summary Statistics

Before digging into the characteristics of the data attributes, we first show a brief summary of statistics (Figure 1), so that we can have the big picture of what we will learn.

	e(sum)	e(mean)	e(sd)	e(min)	e(max)	e(count)
assaults	2245974	4352.663	2120.995	683	8719	516
attend_v	1336.544	2.590201	2.278635	0	12.429	516
attend_m	3776.214	7.318243	4.277742	.35231	24.89	516
attend_n	4619.966	8.953423	3.97134	2.1996	23.858	516
h_chris	4	.0077519	.0877883	0	1	516
h_newyr	4	.0077519	.0877883	0	1	516
h_easter	10	.0193798	.1379897	0	1	516
h_july4	5	.0096899	.0980544	0	1	516
h_mem	12	.0232558	.1508612	0	1	516
h_labor	10	.0193798	.1379897	0	1	516
w_maxa	129.8784	.2517024	.2627543	0	.92228	516
w_maxb	54.64526	.1059017	.1819089	0	.8783	516
w_maxc	3.992	.0077364	.0315128	0	.31658	516
w_mina	10.30059	.0199624	.0625719	0	.48247	516
w_minb	29.85578	.05786	.1122015	0	.628	516
w_minc	90.38868	.1751719	.2129573	0	.87154	516
w_rain	162.871	.3156415	.1645241	0	.79198	516
w_snow	24.16877	.0468387	.0945314	0	.65841	516
year1	51	.0988372	.2987329	0	1	516
year2	52	.1007752	.3013229	0	1	516
year3	52	.1007752	.3013229	0	1	516
year4	52	.1007752	.3013229	0	1	516
year5	53	.1027132	.3038785	0	1	516
year6	52	.1007752	.3013229	0	1	516
year7	52	.1007752	.3013229	0	1	516
year8	52	.1007752	.3013229	0	1	516
year9	52	.1007752	.3013229	0	1	516
year10	48	.0930233	.2907469	0	1	516
month1	44	.0852713	.279556	0	1	516
month2	38	.0736434	.2614432	0	1	516
month3	44	.0852713	.279556	0	1	516
month4	42	.0813953	.2737067	0	1	516
month5	45	.0872093	.2824153	0	1	516
month6	43	.0833333	.2766536	0	1	516
month7	43	.0833333	.2766536	0	1	516
month8	45	.0872093	.2824153	0	1	516
month9	42	.0813953	.2737067	0	1	516
month10	45	.0872093	.2824153	0	1	516
month11	43	.0833333	.2766536	0	1	516
month12	42	.0813953	.2737067	0	1	516

Figure 1: Summary of statistics

From Figure 1, we know, the overall dataset is complete, i.e., there's no NaN or Null in this dataset. All the indicator variables, ones with the prefix $h\_$, and the time variables are binary dummy variables. Meantime, the variables with the prefix $w\_$ are attributes denoting the fraction, ranging from zero to one.

Then, we want to show the distribution of the dataset, to see whether there exist any outliers. Our attention will be mainly put on variables of assaults and movie attendance, since they are highly likely to exhibit outliers, while dummy and fraction variables will not.

We draw the box plots to see the outliers. See Figure 2 and Figure 3 for results.

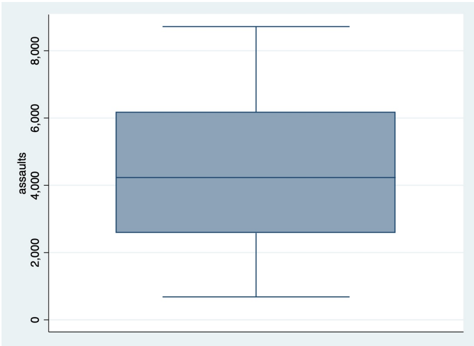

Figure 2: Box plot of assaults

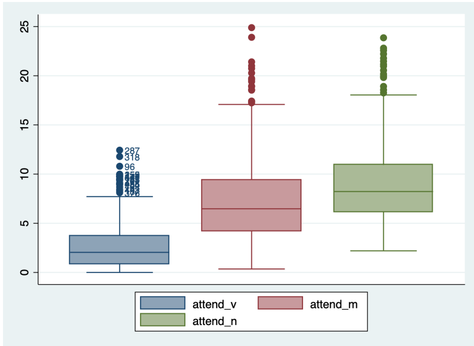

Figure 3: Box plot of attendance

We see, the number of assaults and intimidation does not show any significant outliers. However, for the attendance variables, they all present potential outliers. This may attribute to the quality of the movie. For example, if a movie has high quality, or effective word of mouth, then people will more likely to see the movie, which causes the potential outliers.

2 Regression Analysis

2.1 Regress Number of Assaults on the Year and Month Indicators

First, we run the regression of the logarithm of the number of assaults on the year and month indicators. See Figure 10 in Appendix B.

To determine whether there's a seasonality in assaults, we conduct the test on the month and year indicators.

```
( 1)  month2 = 0
( 2)  month3 = 0
( 3)  month4 = 0
( 4)  month5 = 0
( 5)  month6 = 0
( 6)  month7 = 0
( 7)  month8 = 0
( 8)  month9 = 0
( 9)  month10 = 0
(10)  month11 = 0
(11)  month12 = 0

      F( 11,   495) =   76.26
           Prob > F =   0.0000
```

Figure 4: Month test

```
( 1)  year2 = 0
( 2)  year3 = 0
( 3)  year4 = 0
( 4)  year5 = 0
( 5)  year6 = 0
( 6)  year7 = 0
( 7)  year8 = 0
( 8)  year9 = 0
( 9)  year10 = 0

      F(  9,   495) = 6440.49
           Prob > F =   0.0000
```

Figure 5: Year test

From Figure 4 and Figure 5, we know the time indicators are all significant, in the regression of $\ln(assaults)$. And the overall F-statistics of $month$ indicators is, $F(11, 495) =$

76.26 with a p-value that is essentially zero. Therefore, there exists seasonality within a year.

Taking a step further, we know from the regression coefficient that, there are more assaults in summer than that in winter.

Another finding is that, there's an increasing trend of assaults. This corroborates with the notes in the data document.

2.2 Regress Attendance on the Year and Month Indicators

In this part, we conduct the regression of *attend* on the time indicators. See Figure 11 in Appendix B for the regression results.

Similarly, to test the seasonality, we check the F-statistics of month indicators. See Figure 6 and Figure 7.

```
 ( 1)  month2 = 0
 ( 2)  month3 = 0
 ( 3)  month4 = 0
 ( 4)  month5 = 0
 ( 5)  month6 = 0
 ( 6)  month7 = 0
 ( 7)  month8 = 0
 ( 8)  month9 = 0
 ( 9)  month10 = 0
 (10)  month11 = 0
 (11)  month12 = 0

       F( 11,   495) =   33.72
            Prob > F =    0.0000
```

Figure 6: Month test

```
 ( 1)  year2 = 0
 ( 2)  year3 = 0
 ( 3)  year4 = 0
 ( 4)  year5 = 0
 ( 5)  year6 = 0
 ( 6)  year7 = 0
 ( 7)  year8 = 0
 ( 8)  year9 = 0
 ( 9)  year10 = 0

       F(  9,   495) =    9.24
            Prob > F =    0.0000
```

Figure 7: Year test

All time indicators are statistically significant, and there exists seasonality within a year, since $F(11, 495) = 33.72$ with a zero p-value. Also, attendance is higher in summer relatively.

2.3 Regress Number of Assaults on Attendance

We now run the regression of $\ln(assaults)$ on *attend*,

$$\ln(assaults) = \beta_0 + \beta_1 attend + u \tag{1}$$

See Figure 12 for the regression results.

Intuitively, we can have the following explanations for the omitted variable bias. In section 2.1, time indicators are significant attributes in affecting the number of assaults. Meantime, in section 2.2, there's also a significant effect of time indicators on the number of assaults. Hence, when leaving the time indicators out of the regression model (as we do in Equation 1), then, the omitted effect is absorbed in the error term u.

Mathematically, we revisit this heteroscedasticity problem with rigorous proof.

With all explanatory variables existing, we have,

$$\ln(assaults) = \beta_0 + \beta_1 \times attend + \beta_2 \times time + u \tag{2}$$

where *time* denotes all the time indicators. In Equation 2, we have,

$$Cov(attend, u) = 0 \qquad Cov(time, u) = 0$$

and

$$attend = \gamma \times time + \epsilon$$

Now, we omit the time indicator variable.

$$\ln(assaults) = \beta_0 + \beta_1 \times attend + u \tag{3}$$

where $\frac{\beta_2 \times time}{\gamma}$ will be absorbed into $\beta_1 \times attend$, and $\frac{\beta_2}{\gamma} \cdot \epsilon$ will be absorbed into u.

Therefore,

$$Cov(attend, \epsilon) = Cov(\gamma \times time + \epsilon, \epsilon) \neq 0$$

which shows the existence of omitted variable bias.

2.4 Regress Number of Assaults on Attendance and Time Indicators

We now include *attend* and time indicators as our independent variables, and run the regression again. See Figure 13 for the regression results[1].

To make a comparison between regression Equation (2) and Equation (3), we can observe the R^2 and $RMSE$.

In Equation 3 (omitting variables), we have the regression results in Figure 12. In the results, we know that this regression has low $R^2 = 0.0633$ and high $RMSE = 0.6107$, which means the logarithm of the number of assaults cannot be perfectly interpreted by *attend* only.

When Equation 2, we add the time indicator variables (Figure 13). We can observe that there's an obvious increase in the regression results. R^2 increases to 0.9918, and $RMSE$ decreases to 0.0583.

Therefore, there truly exists the omitted variable bias in Equation 3. This is consistent with the results in the previous section 2.3.

3 Analysis with Different Controls

In this section, the key regression can be expressed as,

$$\ln(assaults) = \beta_0 + \beta_n \times attend\_n + \beta_m \times attend\_m + \beta_v \times attend\_v + \underline{controls} + u \tag{4}$$

In the following subsections, we will add different controls to estimate the effects of attendance.

3.1 Movie Attendance

Here, we run the regression on the overall attendance, adding $h*$ (holidays), $w*$ (weathers), $year*$, and $month$ as control variables.

The reason why I choose these control variables is straightforward. Since our target is to find the effects of attendance on the number of assaults, the weather conditions, holidays, and time indicators will also affect the attendance, as well as the number of assaults. To avoid the omitted variable bias, we add these variables as controls. Doing

[1]Note: $year1$ and $month9$ are omitted due to the collinearity

so will also lead to many negative influences on the regression model, and I will discuss these issues in Section 4.

The regression results can be found in Figure 14. The model has $R^2 = 0.9960$ and $RMSE = 0.0414$, which shows a good interpretability. And the effect of attendance is significant. An increase in attendance of movies is predicted to reduce assaults by 0.27%. The coefficient is statistically significant at the 1% significance level.

3.2 Different Effects of Different Movies

In this section, we want to investigate whether there's a difference between the effect of watching different kinds of movies. Then, we run the regression again. This time, we strictly follow Equation 4, adding the controls in the previous part (Section 3.1).

See Figure 15 for the regression results. From the results, we observe that $R^2 = 0.9960$ and $RMSE = 0.0414$. And then, we want to test whether there are differences between the effects of different movies. See the below figures (Figure 8 and Figure 9) for the testing results.

```
. test attend_v = attend_m

 ( 1)  attend_v - attend_m = 0

       F(  1,   477) =    0.03
            Prob > F =    0.8730

. test attend_v = attend_n

 ( 1)  attend_v - attend_n = 0

       F(  1,   477) =    1.25
            Prob > F =    0.2639

. test attend_m = attend_n

 ( 1)  attend_m - attend_n = 0

       F(  1,   477) =    1.98
            Prob > F =    0.1603
```

Figure 8: Pairwise test

```
. test attend_v attend_m attend_n

 ( 1)  attend_v = 0
 ( 2)  attend_m = 0
 ( 3)  attend_n = 0

       F(  3,   477) =    8.75
            Prob > F =    0.0000

. test attend_v = attend_m = attend_n

 ( 1)  attend_v - attend_m = 0
 ( 2)  attend_v - attend_n = 0

       F(  2,   477) =    1.16
            Prob > F =    0.3133
```

Figure 9: Overall and individual test

From Figure 8, we reject the null hypothesis, i.e., $attend\_v, attend\_m$, and $attend\_n$ are significant in this regression model. However, they do not differ from one another. And from Figure 9, the F-statistic suggests that, the coefficient β_v, β_m, and β_n) are not significantly different from each other.

Lastly, we focus on the effects of watching different kinds of movies.

- An increase in the attendance of watching strongly violent movies (in millions) is predicted to reduce assaults by 0.33%
- An increase in the attendance of watching mildly violent movies (in millions) is predicted to reduce assaults by 0.32%

- An increase in the attendance of watching non-violent movies (in millions) is predicted to reduce assaults by 0.23%

These are the negative effects, and all the effects are statistically significant. Intuitively, these effects are small.

4 Limitations

It's definitely hard to control for all the attributes that affect the number of assaults, and that might be correlated with movie attendance. I will show a theoretical example.

Intuitively, we want to add some instruments that affect assaults only by affecting attendance. However, these variables can not meet this requirement. Weather conditions can affect movie attendance, since people may prefer to stay at home when the weather is bad. Meanwhile, when it's rainy, it might be inconvenient to start an assault. Hence, weather cannot be selected as an instrument variable. Similarly, holidays do not work in this model. Hence, there are no instrumental variables can be selected to control the confounding and measurement error.

As for controls, in my model, I choose $h*$, $w*$, and time indicators. Even though it might cause a multicollinearity problem. However, this is a trade-off between multicollinearity and omitted variable bias. I choose more variables that may lead to multicollinearity, since we use the dataset for the descriptive purpose, instead of prediction. Therefore, multicollinearity is just a feature of this dataset, and we have more space to interpret the results. However, when it comes to the omitted variable bias, we have less space for interpretation.

A Code

1. Below is the code to draw the box plots, in the step of detecting the outliers.

```
%%%% draw the box plot

% generate id variable so that we can check the outliers directly
gen id=_n
% draw the box plot
graph box assaults attend_v attend_m attend_n, mark(1,mlabel(id))
```

2. Here are the codes for regression purposes.

```
gen ln_assaults=ln(assaults)

reg ln_assaults year2 year3 year4 year5 year6 year7 year8 year9 year10 month2
    month3 month4 month5 month6 month7 month8 month9 month10 month11 month12

test month2 month3 month4 month5 month6 month7 month8 month9 month10 month11
    month12
```

Note: The regressions in part (2) are quite similar in coding. Therefore, here I just show one example of regression codes.
3. Regression with instrumental variables

```
% run the iv regression
ivregress 2sls ln_assaults (attend = h*) year* month* w*

ivregress 2sls ln_assaults (attend_v attend_m attend_n = h*) year* month* w*

% test the differences between different movies
test attend_v = attend_m
test attend_v = attend_n
test attend_m = attend_n
test attend_v = attend_m = attend_n
```

B Regression Results

Source	SS	df	MS			
				Number of obs	=	516
				F(20, 495)	=	2947.53
Model	202.988459	20	10.149423	Prob > F	=	0.0000
Residual	1.70446737	495	.003443368	R-squared	=	0.9917
				Adj R-squared	=	0.9913
Total	204.692927	515	.397461994	Root MSE	=	.05868

ln_assaults	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
year2	.6949052	.0115735	60.04	0.000	.672166	.7176444
year3	1.008377	.0115786	87.09	0.000	.9856279	1.031126
year4	1.245359	.0115786	107.56	0.000	1.222609	1.268108
year5	1.414503	.0115227	122.76	0.000	1.391863	1.437142
year6	1.695259	.0115656	146.58	0.000	1.672535	1.717983
year7	1.850901	.0115709	159.96	0.000	1.828166	1.873635
year8	1.901345	.0115735	164.28	0.000	1.878606	1.924084
year9	1.943667	.0115786	167.87	0.000	1.920918	1.966417
year10	2.07023	.0118298	175.00	0.000	2.046987	2.093473
month2	.0168826	.0130119	1.30	0.195	-.0086827	.042448
month3	.0799446	.0125332	6.38	0.000	.0553197	.1045695
month4	.1297383	.0126637	10.24	0.000	.104857	.1546195
month5	.1802473	.0124478	14.48	0.000	.1557903	.2047043
month6	.166302	.0125986	13.20	0.000	.1415487	.1910552
month7	.1738895	.0125872	13.81	0.000	.1491586	.1986204
month8	.1767969	.0124505	14.20	0.000	.1523346	.2012592
month9	.1976758	.0126724	15.60	0.000	.1727774	.2225741
month10	.1417214	.0124423	11.39	0.000	.1172752	.1661675
month11	.0248213	.0125957	1.97	0.049	.0000737	.0495689
month12	-.0054357	.0126757	-0.43	0.668	-.0303405	.0194692
_cons	6.727564	.0119825	561.45	0.000	6.704021	6.751107

Figure 10: Regression Results (ln(assaults) on time indicators)

Source	SS	df	MS			
				Number of obs	=	516
				F(20, 495)	=	22.86
Model	5951.52608	20	297.576304	Prob > F	=	0.0000
Residual	6444.23116	495	13.0186488	R-squared	=	0.4801
				Adj R-squared	=	0.4591
Total	12395.7572	515	24.0694315	Root MSE	=	3.6081

attend	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
year2	.8971667	.7116308	1.26	0.208	-.5010227	2.295356
year3	2.031086	.7119441	2.85	0.005	.6322808	3.429891
year4	3.399118	.7119475	4.77	0.000	2.000307	4.79793
year5	2.951545	.7085114	4.17	0.000	1.559484	4.343605
year6	2.512028	.7111488	3.53	0.000	1.114786	3.90927
year7	3.122399	.7114723	4.39	0.000	1.724521	4.520277
year8	5.041182	.7116308	7.08	0.000	3.642992	6.439371
year9	4.405701	.7119441	6.19	0.000	3.006896	5.804506
year10	3.727569	.7273897	5.12	0.000	2.298417	5.156721
month2	-.6612766	.8000767	-0.83	0.409	-2.233242	.9106885
month3	-2.168523	.7706441	-2.81	0.005	-3.682659	-.6543859
month4	-3.403276	.7786674	-4.37	0.000	-4.933177	-1.873376
month5	.9224337	.7653902	1.21	0.229	-.5813805	2.426248
month6	2.916713	.7746617	3.77	0.000	1.394682	4.438743
month7	5.393685	.7739631	6.97	0.000	3.873027	6.914343
month8	1.212276	.7655568	1.58	0.114	-.2918659	2.716417
month9	-5.425176	.7792023	-6.96	0.000	-6.956128	-3.894225
month10	-3.325857	.7650514	-4.35	0.000	-4.829006	-1.822709
month11	3.879529	.774484	5.01	0.000	2.357848	5.401211
month12	.6773855	.7794064	0.87	0.385	-.8539673	2.208738
_cons	16.03961	.7367841	21.77	0.000	14.592	17.48722

Figure 11: Regression Results (attendance on time indicators)

Source	SS	df	MS			
				Number of obs	=	516
				F(1, 514)	=	34.71
Model	12.9473829	1	12.9473829	Prob > F	=	0.0000
Residual	191.745544	514	.373045805	R-squared	=	0.0633
				Adj R-squared	=	0.0614
Total	204.692927	515	.397461994	Root MSE	=	.61077

ln_assaults	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
attend	.0323187	.0054859	5.89	0.000	.0215413	.0430962
_cons	7.606019	.1069099	71.14	0.000	7.395984	7.816053

Figure 12: Regression Results (ln(assaults) on attendance)

Source	SS	df	MS			
				Number of obs	=	516
				F(21, 494)	=	2841.30
Model	203.012141	21	9.66724483	Prob > F	=	0.0000
Residual	1.68078536	494	.0034024	R-squared	=	0.9918
				Adj R-squared	=	0.9914
Total	204.692927	515	.397461994	Root MSE	=	.05833

ln_assaults	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
attend	-.001917	.0007266	-2.64	0.009	-.0033447	-.0004894
year1	0	(omitted)				
year2	.6966251	.0115229	60.46	0.000	.6739852	.7192649
year3	1.012271	.0116037	87.24	0.000	.989472	1.035069
year4	1.251875	.0117716	106.35	0.000	1.228746	1.275003
year5	1.420161	.011653	121.87	0.000	1.397265	1.443056
year6	1.700075	.0116406	146.05	0.000	1.677203	1.722946
year7	1.856886	.0117235	158.39	0.000	1.833852	1.87992
year8	1.911009	.0120735	158.28	0.000	1.887287	1.934731
year9	1.952113	.0119464	163.41	0.000	1.928641	1.975585
year10	2.077376	.0120671	172.15	0.000	2.053666	2.101085
month1	-.1872757	.0131992	-14.19	0.000	-.2132092	-.1613422
month2	-.1716607	.0135177	-12.70	0.000	-.19822	-.1451014
month3	-.1114881	.0128087	-8.70	0.000	-.1366545	-.0863218
month4	-.0640615	.0128191	-5.00	0.000	-.0892482	-.0388748
month5	-.0052601	.0133455	-0.39	0.694	-.0314811	.0209609
month6	-.0153824	.0140324	-1.10	0.274	-.0429529	.0121882
month7	-.0030464	.0149035	-0.20	0.838	-.0323285	.0262356
month8	-.0081549	.0134198	-0.61	0.544	-.0345217	.018212
month9	0	(omitted)				
month10	-.05193	.0126158	-4.12	0.000	-.0767172	-.0271428
month11	-.1550173	.0143536	-10.80	0.000	-.1832188	-.1268157
month12	-.1914128	.0134818	-14.20	0.000	-.2179015	-.1649241
_cons	6.945587	.0140757	493.45	0.000	6.917932	6.973243

Figure 13: Regression Results (ln(assaults) on attendance and time indicators)

Source	SS	df	MS			
				Number of obs	=	516
				F(36, 479)	=	3304.70
Model	203.872087	36	5.66311351	Prob > F	=	0.0000
Residual	.820840276	479	.001713654	R-squared	=	0.9960
				Adj R-squared	=	0.9957
Total	204.692927	515	.397461994	Root MSE	=	.0414

ln_assaults	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
attend	-.0027506	.0005627	-4.89	0.000	-.0038562	-.0016449
year1	0	(omitted)				
year2	.4106263	.0797911	5.15	0.000	.2538424	.5674102
year3	.4367938	.1589172	2.75	0.006	.1245328	.7490549
year4	.3606433	.2377028	1.52	0.130	-.1064258	.8277124
year5	.2344226	.3168182	0.74	0.460	-.3881026	.8569478
year6	.2433408	.3962782	0.61	0.539	-.5353176	1.021999
year7	.1050697	.4754863	0.22	0.825	-.829227	1.039366
year8	-.1240344	.5546051	-0.22	0.823	-1.213794	.9657251
year9	-.3603368	.6340406	-0.57	0.570	-1.606181	.8855078
year10	-.5262075	.7131322	-0.74	0.461	-1.927462	.8750465
month1	.2042651	.0540427	3.78	0.000	.0980752	.3104551
month2	.1739719	.0479242	3.63	0.000	.0798043	.2681395
month3	.1700145	.0413457	4.11	0.000	.0887731	.2512558
month4	.1363956	.034769	3.92	0.000	.068077	.2047143
month5	.1184478	.0290921	4.07	0.000	.0612839	.1756117
month6	.0559971	.0225296	2.49	0.013	.0117279	.1002662
month7	.0237848	.0179116	1.33	0.185	-.0114102	.0589798
month8	-.0018843	.012524	-0.15	0.880	-.026493	.0227244
month9	0	(omitted)				
month10	-.0127647	.0124234	-1.03	0.305	-.0371758	.0116464
month11	-.075427	.0188974	-3.99	0.000	-.112559	-.0382949
month12	-.0876991	.0248159	-3.53	0.000	-.1364605	-.0389376
h_chris	-.0938104	.0232797	-4.03	0.000	-.1395533	-.0480675
h_newyr	.2409526	.0226788	10.62	0.000	.1963904	.2855149
h_easter	-.0344594	.0143671	-2.40	0.017	-.0626898	-.006229
h_july4	.0455219	.0202387	2.25	0.025	.0057543	.0852896
h_mem	-.0024116	.0149053	-0.16	0.872	-.0316995	.0268763
h_labor	.0231805	.0140603	1.65	0.100	-.004447	.0508081
wkd_ind	.0055334	.0015187	3.64	0.000	.0025492	.0085176
w_maxa	.1108398	.0133199	8.32	0.000	.0846671	.1370125
w_maxb	.1148737	.018254	6.29	0.000	.079006	.1507415
w_maxc	.0446441	.0687422	0.65	0.516	-.0904295	.1797177
w_mina	-.324795	.0389032	-8.35	0.000	-.4012369	-.248353
w_minb	-.169687	.0266688	-6.36	0.000	-.2220892	-.1172848
w_minc	-.1173522	.0165867	-7.08	0.000	-.1499438	-.0847605
w_rain	-.0303584	.0126602	-2.40	0.017	-.0552348	-.005482
w_snow	-.0616496	.0292703	-2.11	0.036	-.1191637	-.0041355
_cons	6.696597	.0579564	115.55	0.000	6.582716	6.810477

Figure 14: Regression Results (ln(assaults) on attend (overall) time holiday weather)

Source	SS	df	MS			
Model	203.876071	38	5.36515976	Number of obs	=	516
Residual	.816856053	477	.001712486	F(38, 477)	=	3132.96
				Prob > F	=	0.0000
Total	204.692927	515	.397461994	R-squared	=	0.9960
				Adj R-squared	=	0.9957
				Root MSE	=	.04138

ln_assaults	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
attend_v	-.0033232	.0009881	-3.36	0.001	-.0052648	-.0013816
attend_m	-.0031784	.0006538	-4.86	0.000	-.004463	-.0018937
attend_n	-.0022588	.0006525	-3.46	0.001	-.0035409	-.0009767
year1	0	(omitted)				
year2	.4159479	.0799183	5.20	0.000	.2589124	.5729833
year3	.447452	.1591424	2.81	0.005	.1347452	.7601589
year4	.3742058	.237995	1.57	0.117	-.0934424	.841854
year5	.252596	.3171829	0.80	0.426	-.3706526	.8758445
year6	.2664714	.396747	0.67	0.502	-.5131165	1.046059
year7	.133272	.4760929	0.28	0.780	-.8022267	1.068771
year8	-.0918953	.5553203	-0.17	0.869	-1.183072	.9992812
year9	-.3249086	.6347423	-0.51	0.609	-1.572145	.9223281
year10	-.4866673	.7139072	-0.68	0.496	-1.889459	.9161244
month1	.1993596	.0541508	3.68	0.000	.092956	.3057631
month2	.169902	.048034	3.54	0.000	.0755176	.2642864
month3	.1668251	.0414239	4.03	0.000	.0854293	.2482209
month4	.1332172	.034842	3.82	0.000	.0647545	.20168
month5	.1182691	.0290847	4.07	0.000	.0611192	.1754191
month6	.0553613	.0225259	2.46	0.014	.0110991	.0996235
month7	.0255759	.0179677	1.42	0.155	-.0097297	.0608815
month8	-.0012688	.0125338	-0.10	0.919	-.025897	.0233594
month9	0	(omitted)				
month10	-.0130983	.0124262	-1.05	0.292	-.0375152	.0113187
month11	-.0776756	.0189501	-4.10	0.000	-.1149116	-.0404396
month12	-.0890026	.0248452	-3.58	0.000	-.1378222	-.0401831
wkd_ind	.0054498	.0015204	3.58	0.000	.0024623	.0084372
w_maxa	.111121	.0133173	8.34	0.000	.0849532	.1372889
w_maxb	.1123997	.0183266	6.13	0.000	.0763888	.1484106
w_maxc	.0363257	.0689403	0.53	0.598	-.0991386	.1717899
w_mina	-.3319249	.0391935	-8.47	0.000	-.4089381	-.2549117
w_minb	-.1701614	.0266659	-6.38	0.000	-.2225585	-.1177642
w_minc	-.1157241	.0166159	-6.96	0.000	-.1483735	-.0830747
w_rain	-.0309206	.0126745	-2.44	0.015	-.0558253	-.0060158
w_snow	-.0570441	.0294157	-1.94	0.053	-.1148446	.0007563
h_chris	-.0953435	.0232958	-4.09	0.000	-.1411186	-.0495685
h_newyr	.238923	.0227101	10.52	0.000	.1942988	.2835473
h_easter	-.0355562	.0143913	-2.47	0.014	-.0638344	-.007278
h_july4	.0461473	.0202514	2.28	0.023	.0063543	.0859403
h_mem	-.002078	.0149069	-0.14	0.889	-.0313694	.0272134
h_labor	.0232826	.0140561	1.66	0.098	-.0043369	.050902
_cons	6.699593	.0579752	115.56	0.000	6.585675	6.813512

Figure 15: Regression Results (ln(assaults) on attend (different) time holiday weather)